

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 10/585,466
Source: IFRP
Date Processed by STIC: 7/21/06

ENTERED



IFWP

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/585,466

DATE: 07/21/2006
TIME: 18:33:54

Input Set : N:\CrF4\07192006\J585466.raw
Output Set: N:\CRF4\07212006\J585466.raw

1 <110> APPLICANT: Avalon Pharmaceuticals
 2 <120> TITLE OF INVENTION: Cancer-Linked Genes as Targets for Chemotherapy
 3 <130> FILE REFERENCE: 689290-272
 4 <140> CURRENT APPLICATION NUMBER: US/10/585,466
 5 <141> CURRENT FILING DATE: 2006-07-05
 6 <150> PRIOR APPLICATION NUMBER: 60/534,419
 7 <151> PRIOR FILING DATE: 2004-01-06
 8 <160> NUMBER OF SEQ ID NOS: 69
 9 <170> SOFTWARE: PatentIn version 3.3
 11 <210> SEQ ID NO: 1
 12 <211> LENGTH: 1181
 13 <212> TYPE: DNA
 14 <213> ORGANISM: Artificial
 15 <220> FEATURE:
 16 <223> OTHER INFORMATION: cDNA or putative protein derived from a cDNA.
 17 <400> SEQUENCE: 1
 18 cacgaggctg gaagggccca cttcacaccc tgggctcgcc ataaagcggc cgccggccgc 60
 19 cggccccccag acgcgcggcc gctgccatgg cccagccct gtgcccggc ctctccgagt 120
 20 cctggatgct ctctgcggcc tggggcccaa ctggcgcc gccgcctcc gacaaggact 180
 21 gcggccgtc cctcgctcg tccccagact catggggcag caccggcagcc gacagccccg 240
 22 tggcgagccc cgcgcggccca ggacccctcc gggacccccc cgcccccctcc gtaggttaggc 300
 23 gcgccgcgcg cagcagccgc ctggcgccg ggcagaggca gagcgcgcgat gagcgggaga 360
 24 aactgcgcat ggcgcacgctg gcccgcgccc tgcaacgagct ggcgcgtt ctaccgcgt 420
 25 ccgtggcgcc cgcggggccag agcctgacca agatcgagac gtcgcgcctg gctatccgct 480
 26 atatcggcca cctgtcgccc gtgtctaggcc tcagcgagga gagtctccag cgccggtgcc 540
 27 ggcagcgcgg tgacgcgggg tccctcgccc gctgcccgt gtgccccgac gactgccccg 600
 28 cgcagatgca gacacggacg caggtcgagg ggcagggca gggggcgcggg ctgggcctgg 660
 29 tatccgcgt ccgcgcgggg gcgtcctggg gatcccccc tcgcctgcggg ggagcccgag 720
 30 ctgcacccga gccgcgcgac ccgcctgcgc ttgcgcgtgc ccggaaaggc 780
 31 aggcgatgga gccaagccca ccgtcccccc tccttcgggg cgacgtgcgt gctctgttgg 840
 32 agacctggat gcccctctcg cctctggagt ggctgcctga ggagcccaag tgacaaggga 900
 33 caactgacgc cgtctctgtg agcaccgagg cttttggcc tcagcacctt cgaagtggtt 960
 34 ctttggcaga ctgccttcc tggaaagaggg cacggcgat cccgacgggg gcattcctgc 1020
 35 gggtgagagc cgtccccacc gcggcgccccc ttctcagccc ctccctccat ggagggaccc 1080
 36 atagggctag acacttttag gcaagcaggaa ggctgcctt aatgtgaatt tatttatttg 1140
 37 tgaataaact gtactggtgt caaaaaaaaaaaa aaaaaaaaaa a 1181
 39 <210> SEQ ID NO: 2
 40 <211> LENGTH: 268
 41 <212> TYPE: PRT
 42 <213> ORGANISM: Artificial
 43 <220> FEATURE:
 44 <223> OTHER INFORMATION: cDNA or putative protein derived from a cDNA.
 45 <400> SEQUENCE: 2

RAW SEQUENCE LISTING

DATE: 07/21/2006

PATENT APPLICATION: US/10/585,466

TIME: 18:33:54

Input Set : N:\Crf4\07192006\J585466.raw

Output Set: N:\CRF4\07212006\J585466.raw

46 Met Ala Gln Pro Leu Cys Pro Pro Leu Ser Glu Ser Trp Met Leu Ser
 47 1 5 10 15
 48 Ala Ala Trp Gly Pro Thr Arg Arg Pro Pro Pro Ser Asp Lys Asp Cys
 49 20 25 30
 50 Gly Arg Ser Leu Val Ser Ser Pro Asp Ser Trp Gly Ser Thr Pro Ala
 51 35 40 45
 52 Asp Ser Pro Val Ala Ser Pro Ala Arg Pro Gly Thr Leu Arg Asp Pro
 53 50 55 60
 54 Arg Ala Pro Ser Val Gly Arg Arg Gly Ala Arg Ser Ser Arg Leu Gly
 55 65 70 75 80
 56 Ser Gly Gln Arg Gln Ser Ala Ser Glu Arg Glu Lys Leu Arg Met Arg
 57 85 90 95
 58 Thr Leu Ala Arg Ala Leu His Glu Leu Arg Arg Phe Leu Pro Pro Ser
 59 100 105 110
 60 Val Ala Pro Ala Gly Gln Ser Leu Thr Lys Ile Glu Thr Leu Arg Leu
 61 115 120 125
 62 Ala Ile Arg Tyr Ile Gly His Leu Ser Ala Val Leu Gly Leu Ser Glu
 63 130 135 140
 64 Glu Ser Leu Gln Arg Arg Cys Arg Gln Arg Gly Asp Ala Gly Ser Pro
 65 145 150 155 160
 66 Arg Gly Cys Pro Leu Cys Pro Asp Asp Cys Pro Ala Gln Met Gln Thr
 67 165 170 175
 68 Arg Thr Gln Ala Glu Gly Gln Gly Arg Gly Leu Gly Leu Val
 69 180 185 190
 70 Ser Ala Val Arg Ala Gly Ala Ser Trp Gly Ser Pro Pro Ala Cys Pro
 71 195 200 205
 72 Gly Ala Arg Ala Ala Pro Glu Pro Arg Asp Pro Pro Ala Leu Phe Ala
 73 210 215 220
 74 Glu Ala Ala Cys Pro Glu Gly Gln Ala Met Glu Pro Ser Pro Pro Ser
 75 225 230 235 240
 76 Pro Leu Leu Pro Gly Asp Val Leu Ala Leu Leu Glu Thr Trp Met Pro
 77 245 250 255
 78 Leu Ser Pro Leu Glu Trp Leu Pro Glu Glu Pro Lys
 79 260 265
 81 <210> SEQ ID NO: 3
 82 <211> LENGTH: 397
 83 <212> TYPE: DNA
 84 <213> ORGANISM: Artificial
 85 <220> FEATURE:
 86 <223> OTHER INFORMATION: cDNA or putative protein derived from a cDNA.
 87 <400> SEQUENCE: 3
 88 tttttttt ttttttacc tttttacata ttttttatt ttcccttcagc aacatccca 60
 89 atccttccaa gaagagagtt gttgggaggc ctcaggctcg ggccttctc agtcctggc 120
 90 tctgcctggc tgctctgtgc tctgtgtcct ctcccttctt tcgcttcctc caaacattgc 180
 91 tccttcaatc ccaggagaag tctcctcgga tgcgcgc tctaaagcag cccaaggctt 240
 92 gcctcaattg catggtttcc cgagtccctca gctccagaag accaggcaga tgggtggacc 300
 93 ggtgagcagc agggcagccc ctgtgcctct gtctctgcg agtcaactccg aagccggca 360
 94 ggcagcggagg aggaggaggat ttctccaagg acagaag 397
 96 <210> SEQ ID NO: 4

RAW SEQUENCE LISTING

DATE: 07/21/2006

PATENT APPLICATION: US/10/585,466

TIME: 18:33:54

Input Set : N:\Crf4\07192006\J585466.raw
 Output Set: N:\CRF4\07212006\J585466.raw

```

97 <211> LENGTH: 1693
98 <212> TYPE: DNA
99 <213> ORGANISM: Artificial
100 <220> FEATURE:
101 <223> OTHER INFORMATION: cDNA or putative protein derived from a cDNA.
102 <400> SEQUENCE: 4
103 tatccccatc cttagcaggg aggtgctgg catgtgaccc gatgttggaa ttgacaagct      60
104 gctagctagt ccgggccttt tccccccccc tttcctttt ttttttcct cctctccccc 120
105 ccctcccgcc ttcccttctt ttagccacc tcaggggaag caacagatcg tcactcggtg 180
106 ttctcaccga aagcacgtaa tcgcccgtgt aactcatgtt gcgtgggggg cctcccgccg 240
107 cgccgaggaga ggctggggtg cgcccccattt cagcatgctt gtgctcaatt gcaggggtcct 300
108 cgttctcgag tgtgcagagg ggggtgagag ctcaactctc gtcccccacctt cccacccgca 360
109 gctcccccggg tgggtgaggg atgcccctgga ctggggatag ccaggtggga gtccgtcgct 420
110 gtgtggcctg tggctctcga gtctgttctc ctggagtctc gcatttgac ccccttcttc 480
111 gcagtcccccc tcccatagac ttgctctggg aagcgcctct gcctccgacc ctagccggaa 540
112 ccccttcggg gccagagttt gaagccgtgg atgtgcctgc ctgggtggctt gtccgatttg 600
113 cacgggtact tgattacact ctctcattca tggtaacttc cgaagcgtt tagtgccttc 660
114 cggtccctaaa ccgcctacag ccagaacggc ttctccccgc gtttgcac tgatccgcag 720
115 ggcggaaag ggccttcgtc ttaccgggg tccacccctc ccctcatctt ccctgcctac 780
116 ctcttcatcc caccccttgt ccttggagaa actccctctt cctcgctgcc tgccgggctt 840
117 cggagtgact cggcagagac agaggcacag gggctgcct gctgtcacc ggtccaccca 900
118 tctgccttgtt ctgtggagc tgaggactcg ggaaaccatg caattgaggc aagccttggg 960
119 ctgttttaga ggccgtgaca tccgaggaga cttctcctgg gaggtccaaac agccgagctt 1020
120 agcccacccgg gctctggaa agacccgact gaggctaaag ccgcggggaa aggccaagtc 1080
121 cgagttccat ttcttgaaga ggccggcgcg cgttaaggctg tgacattggc cctggcgact 1140
122 ggcttcccag gagctttct ttctcaggag ctccacagcg cggggcatct ccagaaaaact 1200
123 gtcttcagag tgtatttcct ttatcgtca acccagagcc ccaccggcggc taatgcaaga 1260
124 ggcaaaaaaaaa tggttggagg aagaaaaaaca aaggcagggaa gtggcgggg cctgacggtg 1320
125 cgtgtgtgtc tgcagagaag ggagggagcc ggctcagtct ttcttgcattt ttccaaactt 1380
126 caaggtccag gcagccctct gcagggccgg gccccattgc tccccggcgcg gcattggagg 1440
127 tggccggccgg agaggagaag gccaacgcct ggcggcaggct tgcaggcgg aaacggctaa 1500
128 caaggagatt tggtcagcaa aacagaccca gccttccga gccttcgtct gacttggccc 1560
129 gaaagggtgg ggaggggggg ctgcgcaga gcctcaggaa cccttccttc tggggactac 1620
130 catccctgag ctttacgctt cttccacag cctttgcagg cgaaatatcg gaataaagtg 1680
131 ggtccaggcgc cca 1693
133 <210> SEQ ID NO: 5
134 <211> LENGTH: 76
135 <212> TYPE: PRT
136 <213> ORGANISM: Artificial
137 <220> FEATURE:
138 <223> OTHER INFORMATION: cDNA or putative protein derived from a cDNA.
139 <400> SEQUENCE: 5
140 Met Phe Gly Gly Arg Lys Thr Lys Ala Gly Ser Gly Gly Gly Leu Thr
141 1 5 10 15
142 Val Arg Val Cys Leu Gln Arg Arg Glu Gly Ala Gly Ser Val Ser Ser
143 20 25 30
144 Cys Phe Ser Lys Leu Gln Gly Pro Gly Ser Pro Leu Gln Gly Arg Ala
145 35 40 45
146 Pro Leu Leu Pro Ala Arg His Trp Arg Trp Pro Pro Glu Arg Arg Arg

```

RAW SEQUENCE LISTING

DATE: 07/21/2006

PATENT APPLICATION: US/10/585,466

TIME: 18:33:54

Input Set : N:\Crf4\07192006\J585466.raw
 Output Set: N:\CRF4\07212006\J585466.raw

147	50	55	60				
148	Pro	Thr	Pro	Ala Pro Gly Leu Ser Gly Gly Asn Gly			
149	65	70	75				
151 <210>	SEQ ID NO:	6					
152 <211>	LENGTH:	3611					
153 <212>	TYPE:	DNA					
154 <213>	ORGANISM:	Artificial					
155 <220>	FEATURE:						
156 <223>	OTHER INFORMATION:	cDNA or putative protein derived from a cDNA.					
157 <400>	SEQUENCE:	6					
158	tatccccatc	cttagcaggg	aggtgctgg	catgtgaccc	gatgttggaa	ttgacaagct	60
159	gctagctagt	ccgggcctt	tccccc	tttc	tttttctt	cctctccc	120
160	ccctccggc	ttccttctt	tgtagccacc	tcaggggaag	caacagatcg	tcactcggt	180
161	tttc	aaggcactaa	tcgcccgtgt	aactcatgtt	ggctgggggg	cctccggcg	240
162	cgcgccggaga	ggctgggtg	cgc	catt	gtgctcaatt	gcagggtcct	300
163	cgttctcgag	tgtcagagg	gcccgtgag	ctcaactc	gtccccac	cccacccgca	360
164	gctccccggg	tgggtgaggg	atgccttgg	ctggggatag	ccaggtgg	gtccgtcg	420
165	gtgtggcctg	tgg	tctcgga	gtctgttctc	ctggagtctc	gcatttgac	480
166	gcagtccccc	tcccatagac	ttgctctgg	aagcgcctc	gcctccgacc	ctagccggaa	540
167	cccttcggg	gccagagttt	gaagccgtgg	atgtgcctgc	ctgggtgg	gtccgattt	600
168	cacggta	ct	tgattacact	ctctcattca	tggta	cgaagcgtt	660
169	cgtccctaaa	ccgcctacag	ccagaacggc	ttctcccg	gtttgtcac	tgatccgcag	720
170	ggcccggaa	ggccttcgtc	ttacccggg	tccac	ctc	ccctgcctac	780
171	ctttcatcc	cac	ttctgt	ccttggagaa	actcc	cctcgct	840
172	cgagtgact	cg	ggcagagac	agaggcacag	ggg	gtgtcacc	900
173	tctgccttgt	tttctgg	tgaggactcg	ggaaaccat	caattgagg	aagccttgg	960
174	ctgctttaga	ggc	gtgaca	tccgaggaga	cttctcttgg	gattgaagg	1020
175	gaggaagcga	aagaaagg	aggacacaga	gcacagag	gccaggcaga	gccaggag	1080
176	gagaaggcc	cagac	ctgtccaa	aactctt	ttgg	atgtgtt	1140
177	gctgaaggaa	aataaaaaaa	tatgtaaaaa	gataac	tttttccc	tctccaggaa	1200
178	atagccaaag	ttat	tttacat	atcttgg	gatttag	ataaaact	1260
179	tatta	atgt	tcaacatcg	tttatttatt	tattgt	ctgactgt	1320
180	aacaaatcta	atcg	tgtt	acttgaaa	agaaatatt	tttatgt	1380
181	ataatggc	accatt	gt	at	tttcat	tttttccaa	1440
182	ctctgt	gat	aaact	ttttag	at	tttgcct	1500
183	ccaccacc	ttt	aca	tttac	aca	tttggac	1560
184	aatgaact	ttt	ca	tttacaa	gaa	tttcat	1620
185	ctgaact	ttt	ca	tttcat	gt	tttgg	1680
186	agagggc	at	gt	tttcat	gt	tttgg	1740
187	tagc	ttt	tttcat	tttgg	tttgg	tttgg	1800
188	cttctattt	cag	tttcat	tttgg	tttgg	tttgg	1860
189	ttagagac	aat	tttcat	tttgg	tttgg	tttgg	1920
190	cctggctt	ttt	tttcat	tttgg	tttgg	tttgg	1980
191	cctac	ttt	tttcat	tttgg	tttgg	tttgg	2040
192	ccaccatt	ttt	tttcat	tttgg	tttgg	tttgg	2100
193	taggctt	ttt	tttcat	tttgg	tttgg	tttgg	2160
194	gacagac	ttt	tttcat	tttgg	tttgg	tttgg	2220
195	gggtt	ttt	tttcat	tttgg	tttgg	tttgg	2280
196	cctctt	ttt	tttcat	tttgg	tttgg	tttgg	2340

RAW SEQUENCE LISTING

DATE: 07/21/2006

PATENT APPLICATION: US/10/585,466

TIME: 18:33:54

Input Set : N:\Crf4\07192006\J585466.raw

Output Set: N:\CRF4\07212006\J585466.raw

197	cggtatgggg gtgtgcgctg ttccgcataa ccactgcctc ccatgtcctc ctgcaggggcc	2400
198	gaaccgagag ggtgctggca ggctggatc ccacgggtgt ccgcaggaga caaaggcgaa	2460
199	ttccggagaa aaggctgggg ctgagaaaagg cgctccggga gcccgtggca gggcaattcg	2520
200	gcaggctgca ccgaagccga gtgcccgag ggacttgcgc cccggaaggg ggtgtgtggg	2580
201	ggcgcgtccg tgaagatgga tgagggaaaa gtttttcat atcagcagaa gggaaaacgc	2640
202	ctggagtggc cgaacacttt tagttgcccc gcaggaatag gagacggta ctcagctccc	2700
203	caaggctgca caatatccca gcttgcggc ctcctgcct cgtgttcgaa atatgctggc	2760
204	ggtgtaatg tgaagggttc tccaggcagg cggctgggc gagggggtggg cgcagctctg	2820
205	aggatcagac caaatctagg gaaaaaaggg gaggcagacc ttccggcct tggtttgac	2880
206	tttgctctga gcctatgatt gacttccgc ctttgcagga ggtccaacac ccgagcttag	2940
207	cccacccggc tctggaaaag acccgactga ggctaaagcc gccccggaaag gccaagtccg	3000
208	agttccattt cttgaagagg cccgcgcgc taaggctgt acattggccc tggcactgg	3060
209	cttcccgagga gctgttctt ctcaggagct ccacagcgcg gccatctcc agaaaaactgt	3120
210	cttcagagtg tatttcctt tatcgtaac ccagagcccc accgcggcta atgcaagagg	3180
211	ccaaaaaatg tttggaggaa gaaaaacaaa ggcaggaaatg ggcggcggcc tgacgggtcg	3240
212	tgtgtgtctg cagagaaggg agggagccgg ctcagtctt tcttggggg ccaaacttca	3300
213	aggccctctgc agggccgggc cccattgctc cccgcgcggc attggaggtg	3360
214	gccgcggag aggagaaggc caacgcctgc gccaggctt tcaggcggaa acggctaaca	3420
215	aggagattt gtcagcaaaa cagaccccagc cttcccgagg cttcgtctga cttggcccga	3480
216	aaggttgggg agggggggct tgccgcagac ctcaggacc ttccctctcg gggactacca	3540
217	tccctgagcc ttacgcttct ttccacagcc tttgcaggcg gaatatcgga ataaagtggg	3600
218	tccaggcgcc a	3611
220 <210>	SEQ ID NO: 7	
221 <211>	LENGTH: 82	
222 <212>	TYPE: PRT	
223 <213>	ORGANISM: Artificial	
224 <220>	FEATURE:	
225 <223>	OTHER INFORMATION: cDNA or putative protein derived from a cDNA.	
226 <400>	SEQUENCE: 7	
227	Met Gly Val Cys Ala Val Pro His Asn His Cys Leu Pro Cys Pro Pro	
228	1 5 10 15	
229	Arg Gly Pro Asn Arg Glu Gly Ala Gly Arg Ala Gly Ser His Gly Cys	
230	20 25 30	
231	Pro Gln Glu Thr Lys Ala Asn Ser Gly Glu Lys Ala Gly Ala Glu Lys	
232	35 40 45	
233	Gly Ala Pro Gly Ala Ala Gly Arg Ala Ile Arg Gln Ala Ala Pro Lys	
234	50 55 60	
235	Pro Ser Ala Arg Arg Asp Leu Pro Pro Gly Arg Gly Cys Val Gly Ala	
236	65 70 75 80	
237	Leu Pro	
239 <210>	SEQ ID NO: 8	
240 <211>	LENGTH: 3264	
241 <212>	TYPE: DNA	
242 <213>	ORGANISM: Artificial	
243 <220>	FEATURE:	
244 <223>	OTHER INFORMATION: cDNA or putative protein derived from a cDNA.	
245 <400>	SEQUENCE: 8	
246	tatccccatc ctttagcaggg aggtgctgg catgtgaccc gatgttgaaa ttgacaagct	60
247	gctagctagt ccggccctt tccccccccc ttccctttt ttttttccct cctctccccc	120

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 07/21/2006
PATENT APPLICATION: US/10/585,466 TIME: 18:33:55

Input Set : N:\Crf4\07192006\J585466.raw
Output Set: N:\CRF4\07212006\J585466.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:20; N Pos. 2,6,14,725,782,828

Invalid <213> Response:

Use of "Artificial" only as "<213> Organism" response is incomplete, per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27

Seq#:28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51

Seq#:52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/585,466

DATE: 07/21/2006

TIME: 18:33:55

Input Set : N:\Crf4\07192006\J585466.raw
Output Set: N:\CRF4\07212006\J585466.raw

L:773 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:20 after pos.:0

M:341 Repeated in SeqNo=20